

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 19, 2000, 13:26:54 : Search time 20.68 Seconds
(without alignments)
675.764 Million cell updates/sec

Title: US-09-208-140-4
Perfect score: 3109
Sequence: 1 SMSYSWTGALVTPCAAEQK.....FTAGYSGGDIYHSVSHRPR 590

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A_Geneseq_36.*

rd size: 0

Number of hits that pass the threshold: 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2983	95.9	3011	1 W77397	Hepatitis C virus
2	2983	95.9	3011	1 W98021	Infectious hepatitis
3	2978	95.8	3011	1 W98020	Infectious hepatitis
4	2961	95.2	3011	1 R21519	Compiled HCV seque
5	2961	95.2	3011	1 R90931	Hepatitis C virus
6	2961	95.2	593	1 W14789	Hepatitis C virus
7	2961	95.2	612	1 W14790	Recombinant modifi
8	2961	95.2	612	1 W14793	Modified HCV RNA-d
9	2961	95.2	593	1 W14794	Recombinant modifi
10	2961	95.2	3011	1 W34480	HCV polyprotein. C
11	2961	95.2	3011	1 W40038	HCV polyprotein. I
12	2956	95.1	3011	1 R34621	Hepatitis C virus
13	2956	95.1	612	1 W14791	Modified HCV RNA-d
14	2956	95.1	612	1 W14792	Modified HCV RNA-d
15	2954	95.0	3011	1 R40119	HCV genomic amino
16	2954	95.0	3011	1 R79232	HCV sequence. New
17	2944	94.7	3011	1 W77398	Hepatitis C virus
18	2940	94.6	3011	1 R65995	Hepatitis C virus
19	2932	94.3	3011	1 R40120	HCV genomic amino
20	2924	94.0	3011	1 R22154	NANBV Hutch c59 is
21	2760	88.8	2955	1 R08124	Hepatitis C virus
22	2679	86.2	3010	1 R20091	Non-A, non-B viral
23	2679	86.2	3010	1 R20111	Non-A, non-B viral
24	2679	86.2	2201	1 W01680	HCV NS2-NS5B non-s
25	2679	86.2	591	1 R30616	Polypeptide coded
26	2653	85.3	3010	1 R68864	Hepatitis C virus
27	2646	85.1	3010	1 R68622	HCV protein cleava
28	2644	85.0	3010	1 R68622	Partial HCV non-st
29	2644	85.0	3010	1 W98022	Infectious hepatitis
30	2641	84.9	3010	1 R34468	Encoded by full-le
31	2637	84.8	3011	1 R53417	Blood transmissibl
32	2637	84.8	3010	1 R29533	HCV NS4-NS5 peptid
33	2636	84.8	1411	1 R54099	NANBHV El/E2 prote
34	2635	84.8	3014	1 R35207	Hepatitis C virus
35	2629	84.6	3014	1 R35207	Hepatitis C virus
36	2614	84.1	590	1 W37130	Hepatitis C virus
37	2610	83.9	2354	1 R41435	PT-NANBH virus non
38	2576	82.9	3010	1 R34580	Human hepatitis C
39	2508	80.7	3011	1 R67588	Hepatitis C virus

40 2416 77.7 2261 1 P90164
41 2416 77.7 2436 1 P92050
42 2416 77.7 2462 1 P90288
43 2416 77.7 2772 1 R08123
44 2411 77.5 2894 1 R24440
45 2411 77.5 2894 1 R70230

Peptide encoded by
Sequence encoded i
Peptide encoded by
Hepatitis C virus
Composite HCV HC-J
Composite hepatitis

ALIGNMENTS

RESULT 1

W77397
ID W77397 standard; Protein; 3011 AA.
AC W77397;
DT 11-JAN-1999 (first entry)
DE Hepatitis C virus H77 polyprotein.
KW HCV; therapy; diagnosis; vector; gene therapy; vaccine.
OS Hepatitis C virus isolate H77.
PN W09839031-Al.
PD 11-SEP-1998.
PF 26-FEB-1998; U04428.
PR 04-MAR-1997; US-811566.
PA (UNIW) UNIV WASHINGTON.
PI Kolykhalov AA, Rice CM;
DR N-PSDB; V59361.
PT New Hepatitis C virus nucleic acid clones - comprising a 5'-terminal conserved sequence, an open reading frame encoding functional components and a 3'-terminal conserved sequence
PS Disclosure; Page 104-115; 209pp; English.
CC This is the amino acid sequence of the polyprotein encoded by hepatitis C virus (HCV) isolate H77 (see V59361). Its cleavage products form functional components of HCV virus particles and RNA replication machinery. A genetically engineered HCV nucleic acid clone is claimed that comprises at least a functional portion of the HCV H77 nucleic acid sequence. The invention relates to the determination of functional HCV genomic RNA sequences, to construction of infectious HCV DNA clones, and to the use of the clones, or their derivatives, in therapeutic, vaccine and diagnostic applications. The invention is also directed to HCV vectors, e.g. for gene therapy or gene vaccines. The products and methods can also be used for identifying cell lines or animals that are permissive for infection with HCV, for studying HCV infection, isolating functional components of HCV, and for screening for agents capable of modulating HCV replication in vitro and in vivo.
SQ sequence 3011 AA;

Query Match 95.9%; Score 2983; DB 1; Length 3011;
Best Local Similarity 96.6%; Pred. No. 1.7e-286;
Matches 570; Conservative 0; Mismatches 0; Indels 20; Gaps 1;

OY 1 SMSYSWTGALVTPCAAEQKLPINALNSLLRHHNLVYVTSRACQKLRHHNLVYST 60
DB 2421 SMSYSWTGALVTPCAAEQKLPINALNSL-----LRHHNLVYST 2460
OY 61 TSRSACQKQKVTDFDLQVLDSDYDLKVEKAAKSKVKNLLSVEEACSLTPPHSAKSK 120
DB 2461 TSRSACQKQKVTDFDLQVLDSDYDLKVEKAAKSKVKNLLSVEEACSLTPPHSAKSK 2520
OY 121 FGYGAKDVRCARKAVAHINSVWKDLLEDVTPIDTTIMAKNEVFCVQPEKGRKPARLI 180
DB 2521 FGYGAKDVRCARKAVAHINSVWKDLLEDVTPIDTTIMAKNEVFCVQPEKGRKPARLI 2580
OY 181 VPPDLGVRVCEKALYDVYVKLPLAVMGSSYGFQSPGQVFEFLVQAKSKTTPMGFSYD 240
DB 2581 VPPDLGVRVCEKALYDVYVKLPLAVMGSSYGFQSPGQVFEFLVQAKSKTTPMGFSYD 2640
OY 241 TRCFDSTVTSIRTEEALYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCGRRCR 300
DB 2641 TRCFDSTVTSIRTEEALYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCGRRCR 2700

CC the viruses produced by transfection of a host cell with the
CC infectious nucleic acid sequence. Vaccines comprising one or more
CC polypeptides made from the infectious nucleic acid sequence are
CC used to immunize mammals, especially humans, against hepatitis C.
CC The nucleic acid sequences can also be used to induce protective
CC immunity against the virus. The nucleic acid sequences or their
CC encoded proteases (e.g. NS3 protease) can additionally be used to
CC develop screening assays to identify antiviral agents for HCV.
SQ Sequence 3011 AA;

Query Match 95.8%; Score 2978; DB 1; Length 3011;
Best Local Similarity 96.4%; Pred No. 5.4e-286;
Matches 569; Conservative 0; Mismatches 1; Indels 20; Gaps 1;

QY 1 SMSYSWTGALVTPCAAEQKLPINALSNLSLRRHNLVYSTTSRSACQKRLRHNLVYST 60
DB 2421 SMSYSWTGALVTPCAAEQKLPINALSNLSL-----LRHNLVYST 2460
QY 61 TSSRSACQKQKVTDFRLQVLDHYQDLKEVKAASVKYKANLLSVEEACSLTPPHSAKSK 120
DB 2461 TSSRSACQKQKVTDFRLQVLDHYQDLKEVKAASVKYKANLLSVEEACSLTPPHSAKSK 2520
QY 121 FCGAKDVRCARKAVAHINSYWKDLSDVTPIDTINAKNEVFCVQPEKGRKPARLI 180
DB 2521 FCGAKDVRCARKAVAHINSYWKDLSDVTPIDTINAKNEVFCVQPEKGRKPARLI 2580
QY 181 VFPDLGVRVCKRMALYDVVSKLPLAVMGSSYGFQYSPGQVFEFLVQAKSKTTPMGFSYD 240
DB 2581 VFPDLGVRVCKRMALYDVVSKLPLAVMGSSYGFQYSPGQVFEFLVQAKSKTTPMGFSYD 2640
QY 241 TRCFDSTVTSIRTEEATYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGNCYGRRCR 300
DB 2641 TRCFDSTVTSIRTEEATYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGNCYGRRCR 2700
QY 301 ASGVLTTSCGNTLTCTVYKARAACRAAGLQDCTMLVCGDDLVVVCESAGVQEDAAASLRAFT 360
DB 2701 ASGVLTTSCGNTLTCTVYKARAACRAAGLQDCTMLVCGDDLVVVCESAGVQEDAAASLRAFT 2760
QY 361 EAMTRYAPPDPPPOPEYDLEITSCSSNVSVAHGAGKRVYLYTRDPTTPLARAWEYA 420
DB 2761 EAMTRYAPPDPPPOPEYDLEITSCSSNVSVAHGAGKRVYLYTRDPTTPLARAWEYA 2820
QY 421 RHTPNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQLEQALNCEIYGACYSIEPLDLP 480
DB 2821 RHTPNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQLEQALNCEIYGACYSIEPLDLP 2880
QY 481 PIQRLHGLSFLSHSYSGPEINRVAACRLKGVPLRAWRHARSVRARLLSRGGRAL 540
DB 2881 PIQRLHGLSFLSHSYSGPEINRVAACRLKGVPLRAWRHARSVRARLLSRGGRAL 2940
QY 541 CGKYLNFNAVRTKLLTPIAAAGRLDLSGWFTAGYSGGDIYHSVSHARPR 590
DB 2941 CGKYLNFNAVRTKLLTPIAAAGRLDLSGWFTAGYSGGDIYHSVSHARPR 2990

----- 4
ID R21519 standard; Protein: 3011 AA.
AC R21519;
DE 22-JUN-1992 (first entry)
KW Compiled HCV sequence.
OS HCV1; serum; gtl1.
OS Hepatitis C virus 1.
FT Key Location/Qualifiers
FT misc_difference 9 /label= ARG
FT misc_difference 11 /label= THR
FT misc_difference 176 /label= THR
FT misc_difference 334 /label= VAL

FT misc_difference 603 /label= ILE
FT misc_difference 848 /label= (ASN)
FT misc_difference 1114 /label= SER
FT misc_difference 1117 /label= THR
FT misc_difference 1276 /label= LEU
FT misc_difference 1328 /label= (VAL)
FT misc_difference 1454 /label= TYR
FT misc_difference 1471 /label= (SER)
FT misc_difference 1877 /label= (GLY)
FT misc_difference 1948 /label= (HIS)
FT misc_difference 1949 /label= (CYS)
FT misc_difference 2021 /label= (VAL)
FT misc_difference 2349 /label= (SER)
FT misc_difference 2385 /label= (PHE)
FT misc_difference 2386 /label= (ALA)
FT misc_difference 2502 /label= (PHE)
FT misc_difference 2690 /label= (GLY)
FT misc_difference 2996 /label= (PRO)
FN W09202642-A.
PD 20-FEB-1992.
PF 12-AUG-1991; U05728.
PR 10-AUG-1990; US-566209.
PA (CHIR-) CHIRON CORP.
PI Houghton M, Choo QL, Kuo G, Weiner AJ, Urdea MS, Irvine BD;
PI Kolberg JH;
PI WPI; 92-080094/10.
DR N-PSDB; Q21744.
DR Regents for isolating, amplifying and detecting HCV
PT polynucleotide(s) - used to monitor spread of blood-borne non-a,
PT non-b hepatitis virus infection and screen blood samples for
PT virus
PS Disclosure; Fig 1; 67pp; English.
CC Heterogeneities in cloned DNAs of HCV1 are indicated by the amino
CC acid indicated in the features, the parentheses indicated that the
CC heterogeneity was detected at or near to the 5'- or 3'-end of the
CC HCV in the clone.
CC The sequence is derived from a composite HCV cDNA from HCV1, a
CC prototypic HCV. The DNA sequence is based upon sequence information
CC derived from a no. of HCV cDNA clones, which were isolated from a no.
CC of HCV cDNA libraries, including the "c" library present in lambda
CC gtl1 (ATCC No.40394), and from human serum. The HCV cDNA clones
CC were isolated by methods described in W09014436.
CC The clones from which the sequence was derived are 5'clone32,
CC b114a, 18g, ag30a, CA205a, CA290a, CA218a, p14a, CA167b, CA156e,
CC CA84a, CA59a, K9-1 (also called K9-1), 26j, 13i, 12f, 14i, 11b, 7f,
CC 7e, 8h, 33c, 40b, 37b, 35, 36, 8i, 32, 33b, 25c, 14c, 8f, 33f, 33g,
CC 39c, 35f, 19g, 26g, 15e, b5a, 16jh, 6k and p131jh.
SQ Sequence 3011 AA;

Query Match 95.2%; Score 2961; DB 1; Length 3011;
Best Local Similarity 95.6%; Pred. No. 2.6e-284;
Matches 564; Conservative 5; Mismatches 1; Indels 20; Gaps 1;
QY 1 SMSYSWTGALVTPCAAEQKLPINALSNLSLRRHNLVYSTTSRSACQKRLRHNLVYST 60

```

|||||
Db 2421 SMSYSGALVTPCAAEQKLPINALNSL-----LRHNLVYST 2460
QY 61 TSSACORQKKVTFDRLOVLDSDYQVLEKVKAAASKYKANLLSVEEACSLTPPHSAKSK 120
Db 2461 TSSACORQKKVTFDRLOVLDSDYQVLEKVKAAASKYKANLLSVEEACSLTPPHSAKSK 2520
QY 121 FGYGAKDVRCRKAVAHNSVWKDLLEDSTVPTDITIMAKNEVFCVQPEKGGKRPARI 180
Db 2521 FGYGAKDVRCRKAVAHNSVWKDLLEDSTVPTDITIMAKNEVFCVQPEKGGKRPARI 2580
QY 181 VFPLGVRVCEKMAKYDVVTKPLAVMGSSYGFQYSPQGVVEFLVQAWKSKTTPMGFSYD 240
Db 2581 VFPLGVRVCEKMAKYDVVTKPLAVMGSSYGFQYSPQGVVEFLVQAWKSKTTPMGFSYD 2640
QY 241 TRCFDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGNGCYRRRCR 300
Db 2641 TRCFDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGNGCYRRRCR 2700
QY 301 ASGVLTTSCGNTLTCYIKARACRAAGLDQCTMLVCGDDLVCESAGVQEDAAASLRAFT 360
Db 2701 ASGVLTTSCGNTLTCYIKARACRAAGLDQCTMLVCGDDLVCESAGVQEDAAASLRAFT 2760
QY 361 EAMTRYSAAPPDPPQPEYDLELITSCSSNVSAHDGAGKRVYILTRDPTPLARAANETA 420
Db 2761 EAMTRYSAAPPDPPQPEYDLELITSCSSNVSAHDGAGKRVYILTRDPTPLARAANETA 2820
QY 421 RHTPVNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQLEALNCEIYGACYSIEPLDLP 480
Db 2821 RHTPVNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQLEALNCEIYGACYSIEPLDLP 2880
QY 481 PIQRLHGLSAFSLHSYSPGEINRVAACLKGLGVPPLRAWHRARSVRARLLSRGGRRAI 540
Db 2881 PIQRLHGLSAFSLHSYSPGEINRVAACLKGLGVPPLRAWHRARSVRARLLSRGGRRAI 2940
QY 541 CGKYLEFNWAVRTKLTPIAAGRLDLSGWTAGYSGGDIYHVSHPAR 590
Db 2941 CGKYLEFNWAVRTKLTPIAAGRLDLSGWTAGYSGGDIYHVSHPAR 2990

RESULT 5
ID R90931 standard; Protein; 3011 AA.
AC R90931;
DT 15-MAY-1996 (first entry)
DE Hepatitis C virus polyprotein.
US Non-A non-B hepatitis virus; NANBHV; HCV; antigen; detection;
FH Hepatitis C virus.
FT Key Location/Qualifiers
FT misc_difference 1..122
FT /label= antigen
FT /note= "C22; R90936"
FT misc_difference 199..328
FT /label= antigen
FT /note= "S2; R90935"
FT misc_difference 1192..1457
FT /label= antigen
FT /note= "C33C; R90932"
FT misc_difference 1569..1931
FT /label= antigen
FT /note= "C100; R90933"
FT misc_difference 2034..2464
FT /label= antigen
FT /note= "NS5; R90934"
FT
FN EP-693687-A1.
PD 24-JAN-1996.
PF 03-APR-1991; 114016.
PR 04-APR-1990; US-504352.
PA (CHIR) CHIRON CORP.
PI Choo Q, Houghton M, Kuo G;
PK WPI; 96-117956/13.
LK N-PSDB; T12710.

```

```

PT Combinations of synthetic Hepatitis C Virus antigens - provide more
PT effective diagnosis of Non-A, Non-B Hepatitis
PS Disclosure; Fig 1(A-Y); 53pp; English.
CC The combination comprises an HCV antigen from the C domain (pref.
CC C22 - R90936) and at least one HCV antigen from the NS3 (pref. C33c
CC - R90932), NS4 (pref. C100 - R90933), S (pref. S2 - R90935) or NS5
CC (R90934) domain.
CC The antigens may in the form of a fusion protein, a simple physical
CC mixture, or the individual antigens commonly bound to a solid matrix.
CC They are pref. prepd. by recombinant DNA techniques (primers are
CC given in T12711-T12716), but can be synthesised or isolated from
CC HCV using affinity chromatography.
SQ Sequence 3011 AA;

Query Match 95.2%; Score 2961; DB 1; Length 3011;
Best Local Similarity 95.6%; Pred. No. 2.6e-284;
Matches 564; Conservative 5; Mismatches 1; Indels 20; Gaps 1;

QY 1 SMSYSGALVTPCAAEQKLPINALNSL-----LRHNLVYST 60
Db 2421 SMSYSGALVTPCAAEQKLPINALNSL-----LRHNLVYST 2460
QY 61 TSSACORQKKVTFDRLOVLDSDYQVLEKVKAAASKYKANLLSVEEACSLTPPHSAKSK 120
Db 2461 TSSACORQKKVTFDRLOVLDSDYQVLEKVKAAASKYKANLLSVEEACSLTPPHSAKSK 2520
QY 121 FGYGAKDVRCRKAVAHNSVWKDLLEDSTVPTDITIMAKNEVFCVQPEKGGKRPARI 180
Db 2521 FGYGAKDVRCRKAVAHNSVWKDLLEDSTVPTDITIMAKNEVFCVQPEKGGKRPARI 2580
QY 181 VFPLGVRVCEKMAKYDVVTKPLAVMGSSYGFQYSPQGVVEFLVQAWKSKTTPMGFSYD 240
Db 2581 VFPLGVRVCEKMAKYDVVTKPLAVMGSSYGFQYSPQGVVEFLVQAWKSKTTPMGFSYD 2640
QY 241 TRCFDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGNGCYRRRCR 300
Db 2641 TRCFDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGNGCYRRRCR 2700
QY 301 ASGVLTTSCGNTLTCYIKARACRAAGLDQCTMLVCGDDLVCESAGVQEDAAASLRAFT 360
Db 2701 ASGVLTTSCGNTLTCYIKARACRAAGLDQCTMLVCGDDLVCESAGVQEDAAASLRAFT 2760
QY 361 EAMTRYSAAPPDPPQPEYDLELITSCSSNVSAHDGAGKRVYILTRDPTPLARAANETA 420
Db 2761 EAMTRYSAAPPDPPQPEYDLELITSCSSNVSAHDGAGKRVYILTRDPTPLARAANETA 2820
QY 421 RHTPVNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQLEALNCEIYGACYSIEPLDLP 480
Db 2821 RHTPVNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQLEALNCEIYGACYSIEPLDLP 2880
QY 481 PIQRLHGLSAFSLHSYSPGEINRVAACLKGLGVPPLRAWHRARSVRARLLSRGGRRAI 540
Db 2881 PIQRLHGLSAFSLHSYSPGEINRVAACLKGLGVPPLRAWHRARSVRARLLSRGGRRAI 2940
QY 541 CGKYLEFNWAVRTKLTPIAAGRLDLSGWTAGYSGGDIYHVSHPAR 590
Db 2941 CGKYLEFNWAVRTKLTPIAAGRLDLSGWTAGYSGGDIYHVSHPAR 2990

RESULT 6
W14789
ID W14789 standard; Protein; 593 AA.
AC W14789;
DT 02-JUL-1997 (first entry)
DE Hepatitis C virus RNA-dependent RNA polymerase NS5B.
KW RNA-dependent RNA polymerase; RDRP; HCV; RNA replicase; NS5B;
KW diagnosis; antiviral; virucide.
OS Hepatitis C virus type 1a.
FH Key Location/Qualifiers
FT misc_difference 1..122
FT /note= "amino acids at positions 1 (Met) and 2
FT (Ala) are additional to the SMSY amino

```

FT WO9712033-A1.
 PN 03-APR-1997.
 PD 27-SEP-1996; U15571.
 PF 27-SEP-1995; US-004383.
 PR (UYEM-) UNIV EMORY.
 PA AL RH, Hagedorn CH;
 PI WPI: 97-212894/19.
 DR N-PSDB; T63431.
 DR Nucleic acid encoding modified HCV RNA-dependent RNA polymerase -
 PT used to screen compounds for antiviral activity and for diagnosis of
 PT HCV infection.
 PS disclosure: Page 27-30; 50pp; English.
 CC The 68 kDa non-structural protein 5B (NS5B) (W14789) of hepatitis C
 CC virus (HCV) has RNA-dependent RNA polymerase (RDRP) activity. The
 CC region of the HCV genome designated NS5B was identified as a
 CC protein cleavage product of the HCV polyprotein using a vaccinia
 CC virus expression system, and recombinant NS5B was expressed in
 CC E. coli cells transformed with a vector carrying PCR-amplified
 CC NS5B cDNA (T63431). Recombinant modified HCV RDRP (see also
 CC W14790-94) can be used to screen cpds. for antiviral activity and
 CC to raise antibodies diagnostic of HCV infection.
 SQ Sequence 593 AA;

Query Match 95.2%; Score 2961; DB 1; Length 593;
 Best Local Similarity 95.6%; Pred. No. 1.9e-285;
 Matches 564; Conservative 5; Mismatches 1; Indels 20; Gaps 1;

QY 1 SMSYSWTGALVTPCAAEQKLPINALNSLLRHHNLVYSTSRACQKRLRHHNLVYST 60
 DB 3 SMSYSWTGALVTPCAAEQKLPINALNSL-----LRHHNLVYST 42
 QY 61 TSRSACQKQKVTDFRLQVLDSHYQVLYKEVKAASKVKANLLSVEACSLTPPHSAKS 120
 DB 43 TSRSACQKQKVTDFRLQVLDSHYQVLYKEVKAASKVKANLLSVEACSLTPPHSAKS 102
 QY 121 FGYGKDVRCRKAVAHINSVKNLLEDSTVPTDTIMAKNEVFCVQPKGKRPALRI 180
 DB 103 FGYGKDVRCRKAVAHINSVKNLLEDSTVPTDTIMAKNEVFCVQPKGKRPALRI 162
 QY 181 VFPDLGVRCVKALYDVVSKPLAVNGSSYGFQYSPQRFVFLVQAWKSKTTPMGFSYD 240
 DB 163 VFPDLGVRCVKALYDVVTKPLAVNGSSYGFQYSPQRFVFLVQAWKSKTTPMGFSYD 222
 QY 241 TRCFDSTVTESDITREAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGNGCYRRCR 300
 DB 223 TRCFDSTVTESDITREAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGNGCYRRCR 282
 QY 301 ASGVLITSCGNTLTCTYIKARAACRAAGLQDCTMLVCGDDLVTICESAGVQEDAASLRAFT 360
 DB 283 ASGVLITSCGNTLTCTYIKARAACRAAGLQDCTMLVCGDDLVTICESAGVQEDAASLRAFT 342
 QY 361 EAMTRYSAAPGDPQPPQDEYDELITSCSSNVSAHDGAGKRVYLYTRDPTPLARAWEA 420
 DB 343 EAMTRYSAAPGDPQPPQDEYDELITSCSSNVSAHDGAGKRVYLYTRDPTPLARAWEA 402
 QY 421 RHTPVNSWLGNIIMFAPTLWARMLTMHFFSVLIARDQLEALNCEIYGACYSIEPLDLP 480
 DB 403 RHTPVNSWLGNIIMFAPTLWARMLTMHFFSVLIARDQLEALNCEIYGACYSIEPLDLP 462
 QY 481 PIQRLHGLSAFSLHYSVSGEINRAACRLKGLVPLRAWRHARSVRARLLSRGAAI 540
 DB 463 PIQRLHGLSAFSLHYSVSGEINRAACRLKGLVPLRAWRHARSVRARLLSRGAAI 522
 QY 541 CGKYLEFNWAVRTKLTPTAAAGRLDLSGWFTAGYSGGDIYHVSHPARPR 590
 DB 523 CGKYLEFNWAVRTKLTPTAAAGRLDLSGWFTAGYSGGDIYHVSHPARPR 572

RESULT 7
 W14790
 ID W14790 standard; Protein; 612 AA.

AC W14790;
 DT 02-JUL-1997 (first entry)
 DE Recombinant modified HCV RNA-dependent RNA polymerase NS5B.
 KW RNA-dependent RNA polymerase; RDRP; HCV; RNA replicase; NS5B;
 OS diagnosis; antiviral; virucide; hepatitis C virus.
 FH Synthetic.
 PH Key
 FT Location/Qualifiers
 FT misc_difference 2..21
 FT /note= "amino acids 2-21 can be any amino acid
 FT sequence of 1-20 amino acids"
 FT misc_difference 22
 FT /note= "amino acid 22 is Ser, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 23
 FT /note= "amino acid 23 is Met, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 24
 FT /note= "amino acid 24 is Ser, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 25
 FT /note= "amino acid 25 is Tyr, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 26
 FT /note= "amino acid 26 is Ser, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 27
 FT /note= "amino acid 27 is Trp, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 28
 FT /note= "amino acid 28 is Thr, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 29
 FT /note= "amino acid 29 is Gly or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 30
 FT /note= "amino acid 30 is Ala, or may be deleted or
 FT substd. by His"
 FT region 31..612
 FT /note= "amino acids 31-612 comprise amino acids
 FT 12-593 of RDRP NS5B"
 FT WO9712033-A1.
 PN 03-APR-1997.
 PD 27-SEP-1996; U15571.
 PF 27-SEP-1995; US-004383.
 PR (UYEM-) UNIV EMORY.
 PA AL RH, Hagedorn CH;
 PI WPI: 97-212894/19.
 DR N-PSDB; T63432.
 DR Nucleic acid encoding modified HCV RNA-dependent RNA polymerase -
 PT used to screen compounds for antiviral activity and for diagnosis of
 PT HCV infection.
 PS Claim 7; Refer to Page 38; 50pp; English.
 CC A modified recombinant hepatitis C virus (HCV) RNA-dependent
 CC RNA-polymerase (RDRP) (W14790) includes up to 20 added amino acids
 CC (pref. a single serine residue), and up to 9 deleted or substd.
 CC amino acids at the N-terminus, joined to amino acid residues 12-593
 CC of HCV RDRP NS5B (see also W14789). Preferably, 1-5 amino acids
 CC are substd. by Ala or His (to facilitate purification by Ni
 CC affinity chromatography) and the NS5B portion has R570H, R572H or
 CC W573H substds. (see also W14791-94). The recombinant HCV RDRP can
 CC be expressed in mammalian host cells and used to screen cpds. for
 CC antiviral activity and to raise antibodies diagnostic of HCV
 CC infection.
 SQ Sequence 612 AA;

Query Match 95.2%; Score 2961; DB 1; Length 612;
 Best Local Similarity 95.6%; Pred. No. 2e-285;
 Matches 564; Conservative 5; Mismatches 1; Indels 20; Gaps 1;
 QY 1 SMSYSWTGALVTPCAAEQKLPINALNSLLRHHNLVYSTSRACQKRLRHHNLVYST 60
 DB 22 SMSYSWTGALVTPCAAEQKLPINALNSL-----LRHHNLVYST 61

QY 61 TSRSACORQKVTDFRLQVLDHYQDLVKEYKAAASKYKANLLSVEEACSLTPPHSAKSK 120
 DB 62 TSRSACORQKVTDFRLQVLDHYQDLVKEYKAAASKYKANLLSVEEACSLTPPHSAKSK 121
 QY 121 FGAGKDVRCRKAVAHINSWKDLLEDVTPIDTTIMAKNEVFCVQPEKGRKPARLI 180
 DB 122 FGAGKDVRCRKAVAHINSWKDLLEDVTPIDTTIMAKNEVFCVQPEKGRKPARLI 181
 QY 181 VFPDLGVRVCEKMAKYDVVSKPLAVMGSSYGFQYSPGQVVEFLVQAWKSKTTPMGFSYD 240
 DB 182 VFPDLGVRVCEKMAKYDVVSKPLAVMGSSYGFQYSPGQVVEFLVQAWKSKTTPMGFSYD 241
 QY 241 TRCFDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCYRRCR 300
 DB 242 TRCFDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCYRRCR 301
 QY 301 ASGVLTTSCGNLTTCYIKARAACRAAGLQDCTMLVCGDDLVCESAGVQEDAAASLRAFT 360
 DB 302 ASGVLTTSCGNLTTCYIKARAACRAAGLQDCTMLVCGDDLVCESAGVQEDAAASLRAFT 361
 QY 361 EAMTRYAPPDPPQPEYDLEITSCSSNVSAHVGAGKRVYILTRDPTPLARAAMETA 420
 DB 362 EAMTRYAPPDPPQPEYDLEITSCSSNVSAHVGAGKRVYILTRDPTPLARAAMETA 421
 QY 421 RHTPVNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQLEQALNCEIYGACYSIEPLDLP 480
 DB 422 RHTPVNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQLEQALNCEIYGACYSIEPLDLP 481
 QY 481 PIIQRLHGLSAFSLHSYSPGEINRVAACLRKLGVPPLAWRHARSVRARLLSRGGRRAI 540
 DB 482 PIIQRLHGLSAFSLHSYSPGEINRVAACLRKLGVPPLAWRHARSVRARLLSRGGRRAI 541
 QY 541 CGKYLFNNAVTRTKLTPIAAAGRLDLSGWFTAGYSGGDIYHSVSHARPR 590
 DB 542 CGKYLFNNAVTRTKLTPIAAAGRLDLSGWFTAGYSGGDIYHSVSHARPR 591

RESULT 8

W14793
 ID W14793 standard; Protein: 612 AA.
 AC W14793;
 DT 02-JUL-1997 (first entry)
 DE Modified HCV RNA-dependent RNA polymerase NS5B (W573H).
 KW RNA-dependent RNA polymerase; RDRP; HCV; RNA replicase; NS5B;
 KW diagnosis; antiviral; virucide; hepatitis C virus.
 OS Synthetic.
 Key Location/Qualifiers
 misc_difference 2..21
 /note= "amino acids 2-21 can be any amino acid"
 /note= "sequence of 1-20 amino acids"
 FT misc_difference 22
 FT /note= "amino acid 22 is Ser, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 23
 FT /note= "amino acid 23 is Met, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 24
 FT /note= "amino acid 24 is Ser, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 25
 FT /note= "amino acid 25 is Tyr, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 26
 FT /note= "amino acid 26 is Ser, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 27
 FT /note= "amino acid 27 is Trp, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 28
 FT /note= "amino acid 28 is Thr, or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 29

FT /note= "amino acid 29 is Gly or may be deleted or
 FT substd. by Ala or His"
 FT misc_difference 30
 FT /note= "amino acid 30 is Ala, or may be deleted or
 FT substd. by His"
 FT 31..590
 FT /note= "amino acids 31-590 comprise amino acids
 FT 12-593 of RDRP NS5B"
 PW K09712033-AL..
 PD 03-APR-1997.
 PF 27-SEP-1996; U15571.
 PR 27-SEP-1995; US-004383.
 PA (UDEM-) UNIV EMORY.
 PI AL RH. Hagedorn CH;
 DR WFI; 97-212894/19.
 PT Nucleic acid encoding modified HCV RNA-dependent RNA polymerase -
 PT used to screen compounds for antiviral activity and for diagnosis of
 PT HCV infection
 PS Claim 11: Refer to Page 38; 50pp; English.
 CC A modified recombinant hepatitis C virus (HCV) RNA-dependent
 CC RNA-polymerase (RDRP) (W14793) includes up to 20 added amino acids
 CC (pref. a single serine residue) and up to 9 deleted or substd.
 CC amino acids at the N-terminus, joined to amino acid residues 12-593
 CC of HCV RDRP NS5B (see also W14789) in which the Trp573 residue is
 CC substd. by His. Preferably, 1-5 amino acids are substd. by Ala or
 CC His to facilitate purification by Ni affinity chromatography. The
 CC recombinant HCV RDRP can be expressed in mammalian host cells and
 CC used to screen cpds. for antiviral activity and to raise antibodies
 CC diagnostic of HCV infection.
 SQ Sequence 612 AA;

Query Match 95.2%; Score 2961; DB 1; Length 612;

Best Local Similarity 95.6%; Pred. No. 2e-285; Matches 564; Conservative 5; Mismatches 1; Indels 20; Gaps 1;

QY 1 SMSYSWTGALVTPCAAEQKLPINALNSLRHNLVYSTTSRSACORQKLRHNLVYST 60
 DB 22 SMSYSWTGALVTPCAAEQKLPINALNSL-----LRHNLVYST 61
 QY 61 TSRSACORQKVTDFRLQVLDHYQDLVKEYKAAASKYKANLLSVEEACSLTPPHSAKSK 120
 DB 62 TSRSACORQKVTDFRLQVLDHYQDLVKEYKAAASKYKANLLSVEEACSLTPPHSAKSK 121
 QY 121 FGAGKDVRCRKAVAHINSWKDLLEDVTPIDTTIMAKNEVFCVQPEKGRKPARLI 180
 DB 122 FGAGKDVRCRKAVAHINSWKDLLEDVTPIDTTIMAKNEVFCVQPEKGRKPARLI 181
 QY 181 VFPDLGVRVCEKMAKYDVVSKPLAVMGSSYGFQYSPGQVVEFLVQAWKSKTTPMGFSYD 240
 DB 182 VFPDLGVRVCEKMAKYDVVSKPLAVMGSSYGFQYSPGQVVEFLVQAWKSKTTPMGFSYD 241
 QY 241 TRCFDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCYRRCR 300
 DB 242 TRCFDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCYRRCR 301
 QY 301 ASGVLTTSCGNLTTCYIKARAACRAAGLQDCTMLVCGDDLVCESAGVQEDAAASLRAFT 360
 DB 302 ASGVLTTSCGNLTTCYIKARAACRAAGLQDCTMLVCGDDLVCESAGVQEDAAASLRAFT 361
 QY 361 EAMTRYAPPDPPQPEYDLEITSCSSNVSAHVGAGKRVYILTRDPTPLARAAMETA 420
 DB 362 EAMTRYAPPDPPQPEYDLEITSCSSNVSAHVGAGKRVYILTRDPTPLARAAMETA 421
 QY 421 RHTPVNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQLEQALNCEIYGACYSIEPLDLP 480
 DB 422 RHTPVNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQLEQALNCEIYGACYSIEPLDLP 481
 QY 481 PIIQRLHGLSAFSLHSYSPGEINRVAACLRKLGVPPLAWRHARSVRARLLSRGGRRAI 540
 DB 482 PIIQRLHGLSAFSLHSYSPGEINRVAACLRKLGVPPLAWRHARSVRARLLSRGGRRAI 541
 QY 541 CGKYLFNNAVTRTKLTPIAAAGRLDLSGWFTAGYSGGDIYHSVSHARPR 590

```

Db 542 CGKLFNNAVRTKLTPIAAGQLDLSGWFTAGYSGGDIYHSVSHARPR 591
|||||
RESULT 9
W14794
ID W14794 standard; Protein; 593 AA.
AC W14794;
DT 02-JUL-1997 (first entry)
DE Recombinant modified HCV RNA-dependent RNA polymerase NS5B.
KW RNA-dependent RNA polymerase; RDRP; HCV; RNA replicase; NS5B;
KW diagnosis; antiviral; virucide; hepatitis C virus.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_difference 3 /note= "amino acid 22 is Ser, or may be deleted or
FT substd. by Ala or His"
FT misc_difference 4 /note= "amino acid 23 is Met, or may be deleted or
FT substd. by Ala or His"
FT misc_difference 5 /note= "amino acid 5 is Ser, or may be substd. by
FT Ala or His"
FT misc_difference 6 /note= "amino acid 6 is Tyr, or may be substd. by
FT Ala or His"
FT misc_difference 7 /note= "amino acid 7 is Ser, or may be substd. by
FT Ala or His"
FT misc_difference 8 /note= "amino acid 8 is Trp, or may be substd. by
FT Ala or His"
FT misc_difference 9 /note= "amino acid 9 is Thr, or may be substd. by
FT Ala or His"
FT misc_difference 10 /note= "amino acid 10 is Gly or may be substd. by
FT Ala or His"
FT misc_difference 11 /note= "amino acid 11 is Ala, or may be substd. by
FT His"
FT region 12..593
FT /note= "amino acids 12-593 comprise amino acids
FT 12-593 of RDRP NS5B"
PN W09712033-A1.
PD 03-APR-1997.
PF 27-SEP-1996; U15571.
PF 27-SEP-1995; US-004383.
PR (UYEM-) UNIV EMORY.
AL RH, Hagedorn CH;
WPI: 97-212894/19
DR Nucleic acid encoding modified HCV RNA-dependent RNA polymerase -
PT used to screen compounds for antiviral activity and for diagnosis of
PT HCV infection
PS Claim 12: Refer to Page 38; 50pp; English.
CC A modified recombinant hepatitis C virus (HCV) RNA-dependent
CC RNA-polymerase (RDRP) (W14790) has up to 9 deleted or substd.
CC amino acids at the N-terminus, joined to amino acid residues 12-593
CC of HCV RDRP NS5B (see also W14789). Preferably, 1-5 amino acids
CC are substd. by Ala or His (to facilitate purification by Ni
CC affinity chromatography) and the NS5B portion may have R570H, R572H
CC or W573H substds. (see also W14791-93). The recombinant HCV RDRP
CC can be expressed in mammalian host cells and used to screen cpds.
CC for antiviral activity and to raise antibodies diagnostic of HCV
CC infection.
SQ Sequence 593 AA;

Query Match 95.2%; Score 2961; DB 1; Length 593;
Best Local Similarity 95.6%; Pred. No. 1.9e-285;
Matches 564; Conservative 5; Mismatches 1; Indels 20; Gaps 1;

QY 1 SMSWSTGALVTPCAAEQKLPINALNSLRLHNLVYSTSRACORQLRHHNLVYST 60
|||||
Db 3 SMSWSTGALVTPCAAEQKLPINALNSLRLHNLVYST 42
|||||
61 TSRSACORQKVTFDRLQVLDLSDHYQDLVKEVKAASAKVKKANLLSVEEACSLTPPHSAKSK 120
|||||
43 TSRSACORQKVTFDRLQVLDLSDHYQDLVKEVKAASAKVKKANLLSVEEACSLTPPHSAKSK 102
|||||
121 FGVGAKDVRCARKAVAHINSVWKKLLEDSVTPIDTTIMAKNEVFCVQPEKGGKRPARLI 180
|||||
103 FGVGAKDVRCARKAVAHINSVWKKLLEDSVTPIDTTIMAKNEVFCVQPEKGGKRPARLI 162
|||||
181 VFPLDGVRCVKALYDVSKLPLAVNGSSYGFQYSPQGVFEFLVQAWKSKKTPMGFSYD 240
|||||
163 VFPLDGVRCVKALYDVSKLPLAVNGSSYGFQYSPQGVFEFLVQAWKSKKTPMGFSYD 222
|||||
241 TRCFDSTVTESDIRTEAIIQCCDLDPQARVAIKSLTERLYVGGPLTNSGKNGCYRRRCR 300
|||||
223 TRCFDSTVTESDIRTEAIIQCCDLDPQARVAIKSLTERLYVGGPLTNSGKNGCYRRRCR 282
|||||
301 ASGVLTTSCGNTLTCTYIKARAACRAAGLQDCTMLVCGDDLWVICESAGVQEDAAASLRAFT 360
|||||
283 ASGVLTTSCGNTLTCTYIKARAACRAAGLQDCTMLVCGDDLWVICESAGVQEDAAASLRAFT 342
|||||
361 EAMTRYSAAPPDPPQPEYDLELITSCSSNVSAHDGAGKRVYILTRDPTTPLARAAWETA 420
|||||
343 EAMTRYSAAPPDPPQPEYDLELITSCSSNVSAHDGAGKRVYILTRDPTTPLARAAWETA 402
|||||
421 RHTPVNSWLNIIIMFAPTLWARMILMTHFFSVLTARDQLEALNCEIYGACYSIEPDLDP 480
|||||
403 RHTPVNSWLNIIIMFAPTLWARMILMTHFFSVLTARDQLEALNCEIYGACYSIEPDLDP 462
|||||
481 PIQRLHGLSAFSLHSPGCEINRVAACLRKLGVPPLRAWHRARSVRARLLARGGAAI 540
|||||
463 PIQRLHGLSAFSLHSPGCEINRVAACLRKLGVPPLRAWHRARSVRARLLARGGAAI 522
|||||
541 CGKLFNNAVRTKLTPIAAGQLDLSGWFTAGYSGGDIYHSVSHARPR 590
|||||
523 CGKLFNNAVRTKLTPIAAGQLDLSGWFTAGYSGGDIYHSVSHARPR 572
|||||

RESULT 10
W34480
ID W34480 standard; Protein; 3011 AA.
AC W34480;
DT 16-MAR-1998 (first entry)
DE HCV polyprotein.
KW PCR primer; amplify; HCV; hepatitis C virus; antigen combination; NS3;
KW C domain; S domain; NS5; HCV polyprotein; anti-HCV antibody; detection;
KW NS4.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT Misc_difference 366 /note= "can optionally be Arg"
FT Misc_difference 372 /note= "can optionally be Thr"
FT Misc_difference 867 /note= "can optionally be Thr"
FT Misc_difference 1341 /note= "can optionally be Val"
FT Misc_difference 2148 /note= "can optionally be Ile"
FT Misc_difference 2883 /note= "can optionally be Asn"
FT Misc_difference 3681 /note= "can optionally be Ser"
FT Misc_difference 3690 /note= "can optionally be Thr"
FT Misc_difference 4167 /note= "can optionally be Leu"
FT Misc_difference 4323 /note= "can optionally be Val"
FT Misc_difference 4701 /note= "can optionally be Tyr"

```

FT Misc_difference 4752 /note= "can optionally be Ser"
 ff Misc_difference 5970 /note= "can optionally be Gly"
 fi Misc_difference 6183 /note= "can optionally be His"
 FT Misc_difference 6186 /note= "can optionally be Cys"
 FT Misc_difference 6402 /note= "can optionally be Val"
 FT Misc_difference 7386 /note= "can optionally be Ser"
 FT Misc_difference 7494 /note= "can optionally be Phe"
 FT Misc_difference 7497 /note= "can optionally be Ala"
 FT Misc_difference 7845 /note= "can optionally be Phe"
 FT Misc_difference 8409 /note= "can optionally be Gly"
 FT Misc_difference 9102 /note= "can optionally be Gly"
 FT Misc_difference 9327 /note= "can optionally be Pro"
 PN US5683864-A.
 PD 04-NOV-1997.
 PF 07-JUL-1992; 910760.
 PR 07-JUL-1992; US-910760.
 PR 18-NOV-1987; US-122714.
 PR 30-DEC-1987; US-139886.
 PR 26-FEB-1988; US-161072.
 PR 06-MAY-1988; US-191263.
 PR 26-OCT-1988; US-263584.
 PR 14-NOV-1988; US-271450.
 PR 17-MAR-1989; US-325338.
 PR 20-APR-1989; US-341334.
 PR 21-APR-1989; US-353896.
 PR 04-APR-1990; US-504352.
 PA (CHIR) CHIRON CORP.
 PI Choo Q, Houghton M, Kuo G;
 DR WPI; 97-548976/50.
 N-PSDB; T99981.
 PT Combination of three hepatitis C virus antigens - used for detection
 of specific antibodies to diagnose infection
 PS Disclosure; Column 23-46; 57pp; English.
 CC This sequence represents the Hepatitis C virus polypeptide. Fragments of
 the DNA encoding this sequence can be amplified and used in the
 combination of HCV antigens of the invention. The HCV antigen combination
 comprises an antigen (Ag1) comprising the C domain (i.e. amino acids (aa)
 1-120 of the HCV polypeptide), or its immunologically reactive fragment
 containing at least 8 aa. It also comprises two additional antigens from
 two different polypeptide domains, including at least 8 aa from the NS3,
 NS4, S or NS5 domains of the polypeptide, corresponding, respectively, to
 aa 1050-1640; 1640-2000; 120-400 and 2000-3011 of the HCV polypeptide.
 CC Alternatively, Ag1 contains at least 8 aa from the 1-122 or 9-177 aa
 regions of the HCV polypeptide. These antigen combinations are used
 diagnostically to detect anti-HCV antibodies, using any standard
 immunoassay format. These antigen combinations have a broader range of
 reactivity with antibodies than any antigen individually.
 SQ Sequence 3011 AA;

Query Match 95.2%; Score 2961; DB 1; Length 3011;
 Best Local Similarity 95.6%; Pred. No. 2.6e-284;
 Matches 564; Conservative 5; Mismatches 1; Indels 20; Gaps 1;
 QY 1 SMSYSWTGALVTPCAAEQKLPINALNSLRRHNLVYSTSRACOROKLRHNLVYST 60
 Db 2421 SMSYSWTGALVTPCAAEQKLPINALNSL-----LRHNLVYST 2460
 QY 61 TSSACOROKKVTDFRLQVLDLSDHYQVLEKVAASVKVANLLSVEEACSLTPPHSAKSK 120
 Db 2461 TSSACOROKKVTDFRLQVLDLSDHYQVLEKVAASVKVANLLSVEEACSLTPPHSAKSK 2520

QY 121 FGAGKDYRCHARKAVAHINSWKDLEDSTPTIDTTMAKNEVEFCVQPEKGRPARLI 180
 Db 2521 FGAGKDYRCHARKAVAHINSWKDLEDSTPTIDTTMAKNEVEFCVQPEKGRPARLI 2580
 QY 181 VFPDLGVRVCEKMAIDVYVSKPLAVMGSSYGFQYSPQORVEFLVQAWKSKKTPMGFSYD 240
 Db 2581 VFPDLGVRVCEKMAIDVYVSKPLAVMGSSYGFQYSPQORVEFLVQAWKSKKTPMGFSYD 2640
 QY 241 TRCFDSTVTESDIRTEEAIYOCDDLDPOARVAIKSLTERLYVGGPLTNSRGNCYRRCR 300
 Db 2641 TRCFDSTVTESDIRTEEAIYOCDDLDPOARVAIKSLTERLYVGGPLTNSRGNCYRRCR 2700
 QY 301 ASGVLTTSCGNTLCYIKARACRAAGLDCTMLVCGDDLVVICESAGVQEDAAISLRAFT 360
 Db 2701 ASGVLTTSCGNTLCYIKARACRAAGLDCTMLVCGDDLVVICESAGVQEDAAISLRAFT 2760
 QY 361 EAMTRYAPPDPPPOPEYDLELITSCSSNSVVAHDGAGKRVYLYTRDPTPLARAWEA 420
 Db 2761 EAMTRYAPPDPPPOPEYDLELITSCSSNSVVAHDGAGKRVYLYTRDPTPLARAWEA 2820
 QY 421 RHTPVNSWLGNIIMPAPLWARMILMTHFFSVLIARDQLEQALDCEIYGACYSIEPLDLP 480
 Db 2821 RHTPVNSWLGNIIMPAPLWARMILMTHFFSVLIARDQLEQALDCEIYGACYSIEPLDLP 2880
 QY 481 PIQRLHGLSAFSLHSYSPGEINRVAACLRKLGVPPLRAWHRARSVRARLLSRRGRAI 540
 Db 2881 PIQRLHGLSAFSLHSYSPGEINRVAACLRKLGVPPLRAWHRARSVRARLLSRRGRAI 2940
 QY 541 CGKYLFNNAVTRTKLITPIAAGRLDLSGWFTAGYSGGDIYHSVSHARPR 590
 Db 2941 CGKYLFNNAVTRTKLITPIAAGRLDLSGWFTAGYSGGDIYHSVSHARPR 2990
 RESULT 11
 W40038
 ID W40038 standard; Protein; 3011 AA.
 AC W40038;
 DT 26-MAY-1998 (first entry)
 DE HCV polypeptide.
 KW Hepatitis C virus C domain; HCV; C antigen; immunological activity;
 NS3 domain; NS4 domain; S domain; NS5 domain.
 OS Hepatitis C virus.
 PH Key Location/Qualifiers
 FT Domain 1..120
 FT Modified_site 9 /label= C_domain
 FT /note= "As given in the specification this amino acid can also be Arg"
 FT Modified_site 11 /note= "As given in the specification this amino acid can also be Thr"
 FT Domain 120..400 /label= S_domain
 FT Modified_site 174 /note= "As given in the specification this amino acid can also be Thr"
 FT Modified_site 334 /note= "As given in the specification this amino acid can also be Val"
 FT Modified_site 603 /note= "As given in the specification this amino acid can also be Ile"
 FT Modified_site 847 /note= "As given in the specification this amino acid can also be Asn"
 FT Domain 1050..1640 /label= NS3_domain
 FT Modified_site 1114 /note= "As given in the specification this amino acid can also be Ser"
 FT Modified_site 1217 /note= "As given in the specification this amino acid can also be Val"

FT Modified_site 1276 acid can also be Thr
 FT /note= "As given in the specification this amino
 FT acid can also be Leu"
 FT Modified_site 1328
 FT /note= "As given in the specification this amino
 FT acid can also be Val"
 FT Modified_site 1452
 FT /note= "As given in the specification this amino
 FT acid can also be Tyr"
 FT Modified_site 1472
 FT /note= "As given in the specification this amino
 FT acid can also be Ser"
 FT Domain 1640..2000
 FT /label= NS4_domain
 FT Modified_site 1877
 FT /note= "As given in the specification this amino
 FT acid can also be Gly"
 FT Modified_site 1948
 FT /note= "As given in the specification this amino
 FT acid can also be His"
 FT Modified_site 1949
 FT /note= "As given in the specification this amino
 FT acid can also be Cys"
 FT Domain 2000..3011
 FT /label= NS5_domain
 FT Modified_site 2021
 FT /note= "As given in the specification this amino
 FT acid can also be Val"
 FT Modified_site 2348
 FT /note= "As given in the specification this amino
 FT acid can also be Ser"
 FT Modified_site 2385
 FT /note= "As given in the specification this amino
 FT acid can also be Phe"
 FT Modified_site 2386
 FT /note= "As given in the specification this amino
 FT acid can also be Ala"
 FT Modified_site 2502
 FT /note= "As given in the specification this amino
 FT acid can also be Phe"
 FT Modified_site 2690
 FT /note= "As given in the specification this amino
 FT acid can also be Gly"
 FT Modified_site 2921
 FT /note= "As given in the specification this amino
 FT acid can also be Gly"
 FT Modified_site 2996
 FT /note= "As given in the specification this amino
 FT acid can also be Pro"
 PN US5712087-A.
 PD 27-JAN-1998.
 PR 12-MAY-1995; 440519.
 PR 07-JUL-1992; US-910760.
 PR 04-APR-1990; US-504352.
 PR 12-MAY-1995; US-440519.
 PA (CHIR) CHIRON CORP.
 PI Choo Q, Houghton M, Kuo G;
 PI WPI; 98-119973/11.
 DR Immunossays for hepatitis C virus antibodies - using combinations
 PT of antigenic fragments of HCV polyprotein
 PS Disclosure; Fig 1; 59pp; English.
 CC This sequence represents the hepatitis C virus (HCV) polyprotein which
 CC is used in the construction of novel combinations of HCV antigens that
 CC have a broader range of immunological activity than any single HCV
 CC antigen. An example of such an antigen given in this specification
 CC comprises a first antigen containing at least 8 amino acids of the
 CC C domain of the HCV polyprotein and a second antigen comprising at least
 CC 8 amino acids of the NS3 domain, the NS4 domain, the S domain or the NS5
 CC domain of the HCV polyprotein in the form of a fusion protein, a physical
 CC mixture or bound to a solid matrix.
 CC Note: The features given in the specification as represented in the

CC feature table of W40038 differ from the positions indicated in Figure 1.
 SQ Sequence 3011 AA;
 Query Match 95.2%; Score 2961; DB 1; Length 3011;
 Best Local Similarity 95.6%; Pred No. 2.6e-284;
 Matches 564; Conservative 5; Mismatches 1; Indels 20; Gaps 1;
 QY 1 SMSYSWTGALVTPCAAEEOKLPINALSNLSLRHNLVYSTTSRSACQROKLRHNLVYST 60
 DB 2421 SMSYSWTGALVTPCAAEEOKLPINALSNLSL-----LRHNLVYST 2460
 QY 61 TSRSACQROKLVTPDRLQVLDSDHYQDLKEVKAASKVKANLLSVEEACSLTPPHSAKSK 120
 DB 2461 TSRSACQROKLVTPDRLQVLDSDHYQDLKEVKAASKVKANLLSVEEACSLTPPHSAKSK 2520
 QY 121 FGAGKDVRCARKAVAHINSVWKDLLEDSTVPIDTTINAKNEVFCVQPEKGRKPARLI 180
 DB 2521 FGAGKDVRCARKAVAHINSVWKDLLEDSTVPIDTTINAKNEVFCVQPEKGRKPARLI 2580
 QY 181 VFPDLGVRVCEKMAKYDVVSKPLAVMGSSYGFQYSPGQVFEFLVQAMSKKTPMGFSYD 240
 DB 2581 VFPDLGVRVCEKMAKYDVVSKPLAVMGSSYGFQYSPGQVFEFLVQAMSKKTPMGFSYD 2640
 QY 241 TRCFDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCYRRCR 300
 DB 2641 TRCFDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCYRRCR 2700
 QY 301 ASGVLTTSCGNTLTCTYIKARAACRAAGLQDCPMLVCGDDLVCVICSAGVQEDAAASLRAF 360
 DB 2701 ASGVLTTSCGNTLTCTYIKARAACRAAGLQDCPMLVCGDDLVCVICSAGVQEDAAASLRAF 2760
 QY 361 EAMTRYAPPDPPQPEYDLELITSCSSNVSVAHDGAKRVIYLTDRPTPLARAWEA 420
 DB 2761 EAMTRYAPPDPPQPEYDLELITSCSSNVSVAHDGAKRVIYLTDRPTPLARAWEA 2820
 QY 421 RHTPVNSLGNLIMEFAPTLWARMILMTHFFSVLIARDQLEQALNCEIYGACYSIEPLDLP 480
 DB 2821 RHTPVNSLGNLIMEFAPTLWARMILMTHFFSVLIARDQLEQALNCEIYGACYSIEPLDLP 2880
 QY 481 PIQRLHGLSFAFSLHSPGEINRVAACLRKLGVPPLAWRHRARSVRARLLSRGGRAAI 540
 DB 2881 PIQRLHGLSFAFSLHSPGEINRVAACLRKLGVPPLAWRHRARSVRARLLSRGGRAAI 2940
 QY 541 CGKYLEFNWAVRTKLKLTPIAAAGRLDLSGWFTAGYSGGDIYHSVSHARPR 590
 DB 2941 CGKYLEFNWAVRTKLKLTPIAAAGRLDLSGWFTAGYSGGDIYHSVSHARPR 2990
 RESULT 12
 R31621
 ID R31621 standard; protein; 3011 AA.
 AC R31621;
 DT 24-MAY-1993 (first entry)
 DE Hepatitis C virus (HCV) polyprotein.
 KW Hepatitis; liver disease; HCV; monoclonal antibody; epitope;
 KW Immobilised reagent; immunoassay; diagnosis; detection; treatment;
 KW Infection.
 OS Hepatitis C virus type 1.
 PH Key Location/Qualifiers
 FT domain 1..191
 FT /label= C domain
 FT /note= "nucleocapsid protein"
 FT domain 192..383
 FT /label= E1
 FT /note= "virion envelope protein"
 FT domain 384..800
 FT /label= E2/NS1
 FT /note= "possible envelope"
 FT domain 800..1050
 FT /label= NS2
 FT /note= "unknown function"
 FT domain 1050..1650

```
FT /label= NS3
FT /note= "putative protease domain"
FT 1651..2100
FT /label= NS4
FT /note= "unknown function"
FT 2100..3011
FT /label= NS5
FT /note= "polymerase"
FT
PN W09300365-A.
PD 07-JUN-1993.
PF 24-JUN-1992; U05388.
PR 24-JUN-1991; US-722489.
PA (CHIR ) CHIRON CORP.
PI Chien DY Rutter W;
DR WPI; 93-036334/04.
PT This sequence represents the polyprotein of the HCV prototype
PS isolate HCV1. When compared with all known viral sequences, small
CC structural proteins of the flavivirus family, and with the
CC pestiviruses. The domains shown in the features table are however,
CC tentatively assigned. The polyprotein, and epitopes of it are
CC useful for inducing immunological response in a subject against
CC HCV. The presence of Abs against HCV can be detected using an
CC immunoassay.
SQ Sequence 3011 AA;

Query Match 95.1%; Score 2956; DB 1; Length 3011;
Best Local Similarity 95.4%; Pred. No. 8.2e-284;
Matches 563; Conservative 6; Mismatches 1; Indels 20; Gaps 1;

Qv 1 SMSYSWTGALVTPCAAEQKLPINALNSLRRHNLVYSTTSRSACQKRLRHNLVYST 60
Db 2421 SMSYSWTGALVTPCAAEQKLPINALNSL-----LRHHNLVYST 2460
Qy 61 TSSRCQKQKVTFRQLVDSHYQDVLEKVKAAASKYKANLLSVEEACSLTPHSAKS 120
Db 2461 TSSRCQKQKVTFRQLVDSHYQDVLEKVKAAASKYKANLLSVEEACSLTPHSAKS 2520
Qy 121 FGAGKDVRCARKAVAHINSWKDLLEDSTPDTTMAKNEVFCVQPEKGRKPARLI 180
Db 2521 FGAGKDVRCARKAVAHINSWKDLLEDSTPDTTMAKNEVFCVQPEKGRKPARLI 2580
Qy 181 VFPDLGVRVCEKMAIYVSKLPLAVMGSSYGFQYSPQGRVEFLVQAWKSKTTPMGFSYD 240
Db 2581 VFPDLGVRVCEKMAIYDVVTKLPLAVMGSSYGFQYSPQGRVEFLVQAWKSKTTPMGFSYD 2640
Qy 241 TRCFDSTVTESDIRTEEAIIYQCCDDLPQARVAIKSLTERLYVGGPLTNSRGNGCYRRCR 300
Db 2641 TRCFDSTVTESDIRTEEAIIYQCCDDLPQARVAIKSLTERLYVGGPLTNSRGNGCYRRCR 2700
Qy 301 ASGVLTTCGNTLTCTYIKARACRAAGLDCTMLVCGDDLVVICESAGVQEDAAASLRAFT 360
Db 2701 ASGVLTTCGNTLTCTYIKARACRAAGLDCTMLVCGDDLVVICESAGVQEDAAASLRAFT 2760
Qy 361 EAMTRYAPPDPPQPEYDLELITSCSSNVSVAHGAGKRVYILTRDPTPLARAWEA 420
Db 2761 EAMTRYAPPDPPQPEYDLELITSCSSNVSVAHGAGKRVYILTRDPTPLARAWEA 2820
Qy 421 RHTPNVNLGNITMFAPIIWARMLTWHFFSVLIARDQLEALNCEIYGACYSIEPLDLP 480
Db 2821 RHTPNVNLGNITMFAPIIWARMLTWHFFSVLIARDQLEALNCEIYGACYSIEPLDLP 2880
Qy 481 PIQRLHGLSAFSLHSYSPGENRNVAAKRLKGLVPLAWRHARSVRARLLSRGGRRAI 540
Db 2881 PIQRLHGLSAFSLHSYSPGENRNVAAKRLKGLVPLAWRHARSVRARLLSRGGRRAI 2940
Qy 541 CGKYLEFNNAVRTKLLTPTIAAAGRLDLSGWFTAGYSGGDIYHSVSHARPR 590
Db 2941 CGKYLEFNNAVRTKLLTPTIAAAGRLDLSGWFTAGYSGGDIYHSVSHARPR 2990
```

RESULT 13

```
W14791
ID W14791 standard; Protein; 612 AA.
AC W14791;
DT 02-JUL-1997 (first entry)
DE Modified HCV RNA-dependent RNA polymerase NS5B (R570H).
KW RNA-dependent RNA polymerase; RDRP; HCV; RNA replicase; NS5B;
KW diagnosis; antiviral; virucide; hepatitis C virus.
OS Synthetic.
FH Key
FT Location/Qualifiers
FT misc_difference 2..21
FT /note= "amino acids 2-21 can be any amino acid
FT sequence of 1-20 amino acids"
FT misc_difference 22
FT /note= "amino acid 22 is Ser, or may be deleted or
FT substd. by Ala or His"
FT misc_difference 23
FT /note= "amino acid 23 is Met, or may be deleted or
FT substd. by Ala or His"
FT misc_difference 24
FT /note= "amino acid 24 is Ser, or may be deleted or
FT substd. by Ala or His"
FT misc_difference 25
FT /note= "amino acid 25 is Tyr, or may be deleted or
FT substd. by Ala or His"
FT misc_difference 26
FT /note= "amino acid 26 is Ser, or may be deleted or
FT substd. by Ala or His"
FT misc_difference 27
FT /note= "amino acid 27 is Trp, or may be deleted or
FT substd. by Ala or His"
FT misc_difference 28
FT /note= "amino acid 28 is Thr, or may be deleted or
FT substd. by Ala or His"
FT misc_difference 29
FT /note= "amino acid 29 is Gly or may be deleted or
FT substd. by Ala or His"
FT misc_difference 30
FT /note= "amino acid 30 is Ala, or may be deleted or
FT substd. by His"
FT region
FT 31..612
FT /note= "amino acids 31-612 comprise amino acids
FT 12-593 of RDRP NS5B"
PN W09712033-A1.
PD 03-APR-1997.
PF 27-SEP-1996; U15571.
PR 27-SEP-1995; US-004383.
PA (UYEM-) UNIV EMORY.
PI Al RH, Hagedorn CH;
DR WPI; 97-212894/19.
PT Nucleic acid encoding modified HCV RNA-dependent RNA polymerase -
PT used to screen compounds for antiviral activity and for diagnosis of
PT HCV infection
PS Claim 11: Refer to Page 38; 50pp; English.
CC A modified recombinant hepatitis C virus (HCV) RNA-dependent
CC RNA-polymerase (RDRP) (W14791) includes up to 20 added amino acids
CC (pref. a single serine residue) and up to 9 deleted or substd.
CC amino acids at the N-terminus, joined to amino acid residues 12-593
CC of HCV RDRP NS5B (see also W14789) in which the Arg570 residue is
CC substd. by His. Preferably, 1-5 amino acids are substd. by Ala or
CC His to facilitate purification by Ni affinity chromatography. The
CC recombinant HCV RDRP can be expressed in mammalian host cells and
CC used to screen cpds. for antiviral activity and to raise antibodies
CC diagnostic of HCV infection.
SQ Sequence 612 AA;
```

Query Match 95.1%; Score 2956; DB 1; Length 612;
Best Local Similarity 95.4%; Pred. No. 6.4e-285;
Matches 563; Conservative 5; Mismatches 2; Indels 20; Gaps 1;

```
1 SMSYWTGALVTPCAAEQKLPINALNSLLRHHNLVYST 60
|||||
22 SMSYWTGALVTPCAAEQKLPINALNSL-----LRHNLVYST 61
|||||
61 TSSACORQKVTDFRLQVLDSDHYQVLEKVAASKVKANLLSVEEACSLTPPHSAKS 120
|||||
62 TSSACORQKVTDFRLQVLDSDHYQVLEKVAASKVKANLLSVEEACSLTPPHSAKS 121
|||||
121 FGAGKDVRCARKAVAHINSVKDLLEDSTVDITIMAKNEVFCVQPKGKRPALRI 180
|||||
122 FGAGKDVRCARKAVAHINSVKDLLEDSTVDITIMAKNEVFCVQPKGKRPALRI 181
|||||
181 VFPDLGVRVCEKMAVDVYVSKLPLAVMGSSYGFQYSPGQVRFVQVAKSKTTPMGFSYD 240
|||||
182 VFPDLGVRVCEKMAVDVYVSKLPLAVMGSSYGFQYSPGQVRFVQVAKSKTTPMGFSYD 241
|||||
241 TRCFDSTVTESDIRTEEAIYQCCDLDPOQARVAIKSLTERLYVGGPLTNSRGNGCYRRRC 300
|||||
242 TRCFDSTVTESDIRTEEAIYQCCDLDPOQARVAIKSLTERLYVGGPLTNSRGNGCYRRRC 301
|||||
301 ASGVLTTSCGNTLTCTCIKARACRAAGLQDCTMLVCGDDLWICESAGVQEDAAASRAFT 360
|||||
302 ASGVLTTSCGNTLTCTCIKARACRAAGLQDCTMLVCGDDLWICESAGVQEDAAASRAFT 361
|||||
361 EAMTRYSAAPPDPPQPEYDLELITSCSSNVSAHDGAGKRVYLTTRDPTPLARAWEA 420
|||||
362 EAMTRYSAAPPDPPQPEYDLELITSCSSNVSAHDGAGKRVYLTTRDPTPLARAWEA 421
|||||
421 RHTPVNSWLGNIIMFAPTLWARMLTMHFFSVLIARDQLEALNCIEYACYSIEPLDLP 480
|||||
422 RHTPVNSWLGNIIMFAPTLWARMLTMHFFSVLIARDQLEALNCIEYACYSIEPLDLP 481
|||||
481 PIQRLHGLSFAFSLHSYSGPEINRVAAKRLKGLVPPPLRAWRHARSVRARLLARGGRAAI 540
|||||
482 PIQRLHGLSFAFSLHSYSGPEINRVAAKRLKGLVPPPLRAWRHARSVRARLLARGGRAAI 541
|||||
541 CGKYLEFNWVRTKLTPIAAAGLDLSGWFTAGYSGGDIYHVSVAHPR 590
|||||
542 CGKYLEFNWVRTKLTPIAAAGLDLSGWFTAGYSGGDIYHVSVAHPR 591
|||||

RESULT 14
W14792
ID W14792 standard; Protein; 612 AA.
AC W14792;
DT 02-JUL-1997 (first entry)
DE Modified HCV RNA-dependent RNA polymerase NS5B (R572H).
KW RNA-dependent RNA polymerase; RDRP; HCV; RNA replicase; NS5B;
diagnosis; antiviral; virucide; hepatitis C virus.
Synthetic.
FH Key
FT Location/Qualifiers
FT misc_difference 2..21
/Note- "amino acids 2-21 can be any amino acid
sequence of 1-20 amino acids"
FT misc_difference 22
/Note- "amino acid 22 is Ser, or may be deleted or
substd. by Ala or His"
FT misc_difference 23
/Note- "amino acid 23 is Met, or may be deleted or
substd. by Ala or His"
FT misc_difference 24
/Note- "amino acid 24 is Ser, or may be deleted or
substd. by Ala or His"
FT misc_difference 25
/Note- "amino acid 25 is Tyr, or may be deleted or
substd. by Ala or His"
FT misc_difference 26
/Note- "amino acid 26 is Ser, or may be deleted or
substd. by Ala or His"
FT misc_difference 27
/Note- "amino acid 27 is Trp, or may be deleted or
substd. by Ala or His"
FT misc_difference 28
```

```
FT misc_difference 29
/Note- "amino acid 28 is Thr, or may be deleted or
substd. by Ala or His"
FT misc_difference 30
/Note- "amino acid 29 is Gly or may be deleted or
substd. by Ala or His"
FT misc_difference 31
/Note- "amino acid 30 is Ala, or may be deleted or
substd. by His"
FT region
31..612
/Note- "amino acids 31-612 comprise amino acids
12-593 of RDRP NS5B"
FT W09712033-A1.
PD 03-APR-1997.
PF 27-SEP-1996; U15571.
PR 27-SEP-1995; US-004383.
PA (UYEM-) UNIV EMORY.
PI AL RH, Hagedorn CH;
DR WPI; 97-212894/19.
PT Nucleic acid encoding modified HCV RNA-dependent RNA polymerase -
used to screen compounds for antiviral activity and for diagnosis of
HCV infection
PT Claim 11; Refer to Page 38; 50pp; English.
PS A modified recombinant hepatitis C virus (HCV) RNA-dependent
RNA-polymerase (RDRP) (W14792) includes up to 20 added amino acids
(pref. a single serine residue) and up to 9 deleted or substd.
amino acids at the N-terminus, joined to amino acid residues 12-593
of HCV RDRP NS5B (see also W14789) in which the Arg572 residue is
substd. by His. Preferably, 1-5 amino acids are substd. by Ala or
His to facilitate purification by Ni affinity chromatography. The
recombinant HCV RDRP can be expressed in mammalian host cells and
used to screen cpds. for antiviral activity and to raise antibodies
diagnostic of HCV infection.
CC Sequence 612 AA;
SQ
```

Query Match 95.1%; Score 2956; DB 1; Length 612;

Best Local Similarity 95.6%; Pred. No. 6.4e-285;

Matches 563; Conservative 5; Mismatches 1; Indels 20; Gaps 1;

```
QY 1 SMSYWTGALVTPCAAEQKLPINALNSLLRHHNLVYSTTSSACORQKLRHNLVYST 60
|||||
DB 22 SMSYWTGALVTPCAAEQKLPINALNSL-----LRHNLVYST 61
|||||
QY 61 TSSACORQKVTDFRLQVLDSDHYQVLEKVAASKVKANLLSVEEACSLTPPHSAKS 120
|||||
DB 62 TSSACORQKVTDFRLQVLDSDHYQVLEKVAASKVKANLLSVEEACSLTPPHSAKS 121
|||||
QY 121 FGAGKDVRCARKAVAHINSVKDLLEDSTVDITIMAKNEVFCVQPKGKRPALRI 180
|||||
DB 122 FGAGKDVRCARKAVAHINSVKDLLEDSTVDITIMAKNEVFCVQPKGKRPALRI 181
|||||
QY 181 VFPDLGVRVCEKMAVDVYVSKLPLAVMGSSYGFQYSPGQVRFVQVAKSKTTPMGFSYD 240
|||||
DB 182 VFPDLGVRVCEKMAVDVYVSKLPLAVMGSSYGFQYSPGQVRFVQVAKSKTTPMGFSYD 241
|||||
QY 241 TRCFDSTVTESDIRTEEAIYQCCDLDPOQARVAIKSLTERLYVGGPLTNSRGNGCYRRRC 300
|||||
DB 242 TRCFDSTVTESDIRTEEAIYQCCDLDPOQARVAIKSLTERLYVGGPLTNSRGNGCYRRRC 301
|||||
QY 301 ASGVLTTSCGNTLTCTCIKARACRAAGLQDCTMLVCGDDLWICESAGVQEDAAASRAFT 360
|||||
DB 302 ASGVLTTSCGNTLTCTCIKARACRAAGLQDCTMLVCGDDLWICESAGVQEDAAASRAFT 361
|||||
QY 361 EAMTRYSAAPPDPPQPEYDLELITSCSSNVSAHDGAGKRVYLTTRDPTPLARAWEA 420
|||||
DB 362 EAMTRYSAAPPDPPQPEYDLELITSCSSNVSAHDGAGKRVYLTTRDPTPLARAWEA 421
|||||
QY 421 RHTPVNSWLGNIIMFAPTLWARMLTMHFFSVLIARDQLEALNCIEYACYSIEPLDLP 480
|||||
DB 422 RHTPVNSWLGNIIMFAPTLWARMLTMHFFSVLIARDQLEALNCIEYACYSIEPLDLP 481
|||||
QY 481 PIQRLHGLSFAFSLHSYSGPEINRVAAKRLKGLVPPPLRAWRHARSVRARLLARGGRAAI 540
|||||
```

Db 482 PIIORLHGLSAFSLHSYSPGEINRVAACLRKLGVPPLRAWHRRARSVRARLLARGGAAI 541
 Qy 541 CGKYLEFNNAVTRTKLPTIAAAGRLDLSGWFTAGYSGGDIYHVSHPAR 589
 Db 542 CGKYLEFNNAVTRTKLPTIAAAGRLDLSGWFTAGYSGGDIYHVSHPAR 590
 RESULT 15
 R40119
 ID R40119 standard; Protein; 3011 AA.
 AC R40119;
 DI 27-JAN-1994 (first entry)
 HA HCV genomic amino acid sequence isolated from infected chimpanzee CO.
 NC Hepatitis C Virus; Non-A, non-B hepatitis Virus; HCV; NANBHV;
 KW human growth hormone; HGH; secretion signal; fusion protein;
 vaccine.
 OS Hepatitis C Virus.
 PN WO9315193-A.
 IN 05-AUG-1993.
 DT 29-JAN-1993; U00907.
 PA (ABBO) ABBOTT LAB.
 PI Bode SL, Casey JM, Desai SM, Devare SG, Fraill DE;
 PI Yamaguchi J, Zeck BJ;
 DR WPI; 93-258673/32.
 PT New plasmid pHCV-162 is a mammalian expression systems for HCV
 proteins - useful for diagnosing HCV infection and as vaccines
 PT for preventing HCV infection
 PS Example 1; Page 29-39; 100pp; English.
 CC RNA was isolated from the serum of a chimpanzee (designated "CO")
 CC experimentally infected with HCV and cDNA was prepared from it. The
 CC cDNA was PCR amplified using specific primers with sequences based
 CC on the prototype HCV-1 cDNA sequence (GENBANK M62321). Further
 CC amplification using nested primers resulted in 7 adjacent HCV DNA
 CC fragments which could be assembled into a full-length sequence. The
 CC DNA sequence was determined and translated into the genomic amino
 CC acid sequence. Comparison of the CO genomic amino acid sequence
 CC with that from HCV-1 showed 98 amino acid differences.
 SQ Sequence 3011 AA;

Query Match 95.0%; Score 2954; DB 1; Length 3011;
 Best Local Similarity 95.8%; Pred. No. 1.3e-283;
 Matches 565; Conservative 2; Mismatches 3; Indels 20; Gaps 1;
 Qy 1 SMSYSWTGALVTPCAAEQKLPINALSNLSLRHNLVYSTTSRSACQKQLRHNLVYST 60
 2421 SMSYSWTGALVTPCAAEQKLPINALSNLS-----LRHNLVYST 2460
 Qy 61 TSRSACQKQKVTFRDLQVLDLSDHYQDLKEVKAASKYKANLLSVEEACSLTPPHSAKSK 120
 Db 2461 TSRSACQKQKVTFRDLQVLDLSDHYQDLKEVKAASKYKANLLSVEEACSLTPPHSAKSK 2520
 Qy 121 FGYGAKDVCHARKAVAHINSWKDLLEDVTPIDTTIMAKNEVFCVQPEKGRKPARLI 180
 Db 2521 FGYGAKDVCHARKAVAHINSWKDLLEDVTPIDTTIMAKNEVFCVQPEKGRKPARLI 2580
 Qy 181 VFPDLGVRVCEKMAALYDVVYKSLPLAVMGSSYGFQYSPQORVEFLVQAWKSKKTPMGFSYD 240
 Db 2581 VFPDLGVRVCEKMAALYDVVYKSLPLAVMGSSYGFQYSPQORVEFLVQAWKSKKTPMGFSYD 2640
 Qy 241 TRCFDSTVTESDIRTEEAIYQCDDLPQARVAIKSLTERLYVGGPLTNSRGCGYRRCR 300
 Db 2641 TRCFDSTVTESDIRTEEAIYQCDDLPQARVAIKSLTERLYVGGPLTNSRGCGYRRCR 2700
 Qy 301 ASGVLTTSCGNLTCTYIKARAACRAAGLDCTMLVCGDDLVVVCESAGVQEDAAASIRAF 360
 Db 2701 ASGVLTTSCGNLTCTYIKARAACRAAGLDCTMLVCGDDLVVVCESAGVQEDAAASIRAF 2760
 Qy 361 EAMTRYSAPPDPQPEYDLELITSCSSNVSVAHDCAGKRVYLLTRDPTPLARAWEA 420
 Db 2761 EAMTRYSAPPDPQPEYDLELITSCSSNVSVAHDCAGKRVYLLTRDPTPLARAWEA 2820

Qy 421 RHTPVNSWLGNIIMFAPTLNARMILMTHFFSVLTARDQLDQALNCEIYGACYSIEPLDLP 480
 Db 2821 RHTPVNSWLGNIIMFAPTLNARMILMTHFFSVLTARDQFEOALNCEIYGACYSIEPLDLP 2880
 Qy 481 PIIORLHGLSAFSLHSYSPGEINRVAACLRKLGVPPLRAWHRRARSVRARLLSRGGRAAI 540
 Db 2881 PIIORLHGLSAFSLHSYSPGEINRVAACLRKLGVPPLRAWHRRARSVRARLLSRGGRAAI 2940
 Qy 541 CGKYLEFNNAVTRTKLPTIAAAGRLDLSGWFTAGYSGGDIYHVSHPARPR 590
 Db 2941 CGKYLEFNNAVTRTKLPTIAAAGRLDLSGWFTAGYSGGDIYHVSHPARPR 2990

Search completed: January 19, 2000, 13:27:02
 Job time: 999 sec

